|  |  |  |
| --- | --- | --- |
| **Nama:**  **Ibnu Fajar Setiawan**  **NIM: 065002000006** | C:\Users\RPL-SI 02\Pictures\288px-Trisakti_Logo.svg.png | **MODUL 11**  **Nama Dosen:**  **Ir. Teddy Siswanto, MMSi** |
| **Hari/Tanggal:**  **Hari, 15/06/2022** | **Praktikum Data Warehouse** | **Nama Asisten Labratorium:**   1. **Azhar Rizki Zulma 065001900001** 2. **Nadiya Amanda Rizkania 064001900003** |

**Proyek Akhir Praktikum III**

1. **Teori Singkat**

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

**Cleansing Data**

Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

Berikut ini adalah langkah langkah dalam melakukan cleansing data menggunakan Spoon Pentaho:

1. **Data Preparation**

Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.

|  |
| --- |
|  |

1. **Excel Input**

* Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2020.ktr***

|  |
| --- |
|  |

* Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)

|  |
| --- |
|  |

1. **CSV Input**

* Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Oktober 2020.ktr***

|  |
| --- |
|  |

* Berikut ini adalah konfigurasi pada step CSV File Input.

|  |
| --- |
|  |

1. **Data Filtering**

* Step selanjutnya adalah melakukan select value yang berguna untuk memilah kolom pada excel seperti pada gambar berikut. Kemudian Delete kolom No saja, karena nomor tidak diperlukan lalu Klik OK.

|  |
| --- |
|  |

* Kemudian lakukan ascending data pada proses step Sort Rows

|  |
| --- |
|  |

* Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.

|  |
| --- |
|  |

* Berikut List Data Nama Pasar yang mau dibuang (2020)

|  |
| --- |
|  |

1. **Excel Output**

Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama ***Cabai Merah Keriting [Bulan] [Tahun].xls*** seperti ***Cabai Merah Keriting Agustus 2020.xls*** (File & Sheet)

|  |
| --- |
|  |

1. **Alat dan Bahan**

Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

1. **Elemen Kompetensi**
   1. Tugas Akhir Praktikum III – Screenshot Struktur yang sudah berhasil berjalan

1. Cleansing Januari 2020

|  |
| --- |
|  |

2. Cleansing Februari 2020

|  |
| --- |
|  |

3. Cleansing Maret 2020

|  |
| --- |
|  |

4. Cleansing April 2020

|  |
| --- |
|  |

5. Cleansing Mei 2020

|  |
| --- |
|  |

6. Cleansing Juni 2020

|  |
| --- |
|  |

7. Cleansing Juli 2020

|  |
| --- |
|  |

8. Cleansing Agustus 2020

|  |
| --- |
|  |

9. Cleansing September 2020

|  |
| --- |
|  |

10. Cleansing Oktober 2020

|  |
| --- |
|  |

11. Cleansing November 2020

|  |
| --- |
|  |

12. Cleansing Desember 2020

|  |
| --- |
|  |

1. **File Praktikum**

Github Repository:

|  |
| --- |
| https://github.com/IbnuFajar7/Data-Warehouse/tree/main/Prak-11 |

1. **Soal Latihan**

Soal:

1. Apa itu Cleansing Data?
2. Mengapa Data perlu dilakukan Cleansing?

Jawaban:  
1. Data cleansing adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data

set, tabel, dan database yang korup atau tidak akurat.

2. Menghilangkan kesalahan dan inkonsistensi yang muncul saat beberapa data sources

dikumpulkan dalam satu dataset. Meningkatkan efisiensi kerja karena proses ini akan

memudahkan Anda dan tim pengolah data untuk menemukan apa yang dibutuhkan dari data.

1. **Kesimpulan**
   1. Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
   2. Kita dapat mengetahui bahwa kita bisa melakukan cleansing data yang bisa memudahkan kita untuk cek data yang kemungkinan kita salah input, dengan metode ini kita akan bisa lebih mudah tanpa harus cek satu-satu dan memakan waktu yang lebih lama, waktu ini akan membuat kita lebih efisien
2. **Cek List (✓)**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| **Selesai** | **Tidak Selesai** |
| **1.** | Tugas Akhir Praktikum III |  |  |

1. **Formulir Umpan Balik**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| **1.** | Tugas Akhir Praktikum III | 30 Menit | 1 |

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang